U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL Office Note 75-3

TESTING THE LFM FOR POP FORECASTING

Harry R. Glahn and Joseph R. Bocchieri

June 1975

Testing the LFM for PoP Forecasting

by

Harry R. Glahn and Joseph R. Bocchieri

INTRODUCTION

An objective system for nationwide prediction of probability of precipitation (PoP) has been used operationally by the National Weather Service since October 1971, and distribution to local offices started in January 1972. This system is based on the Model Output Statistics (MOS) technique (Glahn and Lowry, 1972) and uses predictions from the National Meteorological Center's (NMC's) primitive equation (PE) (Shuman and Hovermale, 1968) and the Techniques Development Laboratory's (TDL's) trajectory (TJ) (Reap, 1972) models as input. The question has arisen many times as to whether NMC's limited area fine mesh model (LFM) (Howcroft, 1971) would be useful for MOS PoP prediction. Since our collection of LFM predictors in a form suitable for use in the MOS development system did not start until October 1972, we have been unable to test this idea until recently. Other studies (Ronco, 1972 and 1973) have shown that the LFM does contribute considerable information over and above that from the PE and TJ models for a few stations in New Hampshire and Maine.

Considering the existing NMC operational environment, we might want to use the LFM for either of two reasons:

(1) The LFM is run before the PE, and if only LFM predictors were used, or LFM and TJ predictors (the TJ being driven by the LFM) without PE predictors, PoP forecasts would be available earlier than at present, or

(2) LFM predictors, with or without PE and TJ predictors, might produce better forecasts than would PE and TJ predictors.

This paper describes an experiment, started in Spring 1974, designed to investigate the potential of each of the above possibilities.

PROCEDURE

We could think of three reasonable ways to incorporate LFM predictors into predictive regression equations. These are called methods A, B, and C below:

Method A: We could develop regression equations on our relatively small sample of LFM data, with or without PE and TJ predictors.

The various regression equations that were derived and tested are discussed in three groups--operational equations, small sample equations, and large sample equations. These three groups are identified in Tables 1 and 2, which show verification statistics to be discussed later. In each case, we developed a generalized operator equation by region, each limited to 12 terms. The equations used operationally during the 1973-74 winter were derived on a 456-day sample from the winters 1970-71, 71-72, and 72-73. Derivation of the equations is described by Bocchieri (1974) and definition of the 17 winter regions and station locations are shown in Figure 1. The operational equations for summer 1974 (National Weather Service, 1974) were derived on a 430-day sample from the summers of 1971, 1972, and 1973; the 24 regions used are shown in Figure 2. Item 1 in Tables 1 and 2, refers to these operational equations. The small sample equations were derived on the 120-day 1972-73 winter and 72-day 1973 summer samples. The large sample equations were derived on the same sample as were the operational equations. Therefore, the operational equations were also "large sample", but the distinction is made to emphasize that the operational equations were derived prior to the test periods, while the large sample equations were derived later for testing purposes only.

The 100 predictors screened in developing the operational equations for the winter 1973-74 are given in Table 3; the same types of predictors were screened for the summer, except that the binary limits were slightly different. When LFM predictors were used, they were similar to the PE predictors in Table 3. Since we usually keep the total predictors screened to 100 or less to conserve computer time, addition of new predictors necessitates omitting a few old ones. Some experimentation is necessary to determine proper binary limits of new predictors. Elimination of specific predictors is based on their relatively poor performance when included in previous runs.

We wanted to test the performance of the operational equations when LFM and TJ forecasts were used as input rather than PE and TJ forecasts. However, a few PE predictors were for projections greater than 24 hours, and since the LFM does not run beyond 24 hours, these predictors could not be used. Therefore, "modified" operational equations were derived by eliminating the last-selected terms (in the screening procedure) until no terms with projections greater than 24 hours remained. (The coefficients of the remaining terms were, of course, those appropriate to the modified equations. Actually, these "modified" equations were routinely produced at the same time as were the 12-term equations.) Equations for some regions did not have to be modified; others had a few terms removed. These modified equations correspond to Item 2 of Tables 1 and 2.

The same equations were used in Item 3 as in Item 2, but in Item 3 input in making forecasts was from the LFM and TJ rather than from the PE and TJ.

Items 4-12 in Tables 1 and 2 refer to equations developed on the 120-day winter sample and 72 day summer sample, respectively. The PE, LFM, and TJ models were used singly or in combination. When SIN, COS, CONT is was true for large sample equations (Items 1 and 13); PE alone was significantly worse than PE + TJ. The reason for this inconsistency is not known. In the summer, PE + TJ was better than PE alone for both the small and large sample equations.

5. Adding the first harmonic of the day of year and continuous predictors to the usual binaries gave better forecasts in both winter and summer (Items 7 vs 9, 8 vs 10, 11 vs 12). In fact, the combination PE + TJ + LFM + SIN, COS, CONT (Item 12) gave the best results of all combinations in winter; in summer, that combination gave the best results for the small sample equations only.

6. In winter, the small sample combinations that included the LFM gave better results than the operational equations, which were developed on a sample nearly four times as large. However, in summer, all small sample combinations were significantly worse than the operational equations.

We also did a comparative verification by region between the operational PoP equations and the small sample combination of PE + TJ + LFM (Items 1 and 11 respectively in Tables 1 and 2) for both the summer and winter seasons. The regions were the same as those shown in Figures 1 and 2. For each region, we subtracted the P-score obtained for Item 11 from that obtained for Item 1; the results are shown in Figures 3 and 4 for winter and summer respectively. A positive number indicates that the equations incorporating LFM predictors were better. In general, the use of the LFM gives better results for both summer and winter in the upper Mississippi valley, Great Lakes, Applachian Mountains, and interior New England.

CONCLUSIONS AND PLANS

As discussed above, Tables 1 and 2 show some inconsistent results. For instance, the small sample combinations which included the LFM were better than the operational equations in winter but not in summer. This result could be due to the small number of days available to develop the small sample equations. We will, therefore, do further tests in the near future for both summer and winter seasons with larger developmental data samples--we will not use the LFM in operational PoP for summer until that test is completed. However, based on the winter season results obtained so far, we will incorporate the LFM for operational PoP starting in the winter of 1975-76.

We have also concluded that the addition of the SIN and COS of the day of year and continuous predictors to the usual binary predictors will increase the accuracy of PoP forecasts. These new predictors will be added to the summer 1975 operational PoP system.

ACKNOWLEDGEMENTS

We are thankful to Mr. George Hollenbaugh for preparing data tapes used in this study and to Mr. Thomas Bethem for helping make the t-tests. We also appreciate the editorial help given by Mr. Roger Allen.

Table 1. Verification on 79-case independent data sample of forecasts from winter 1973-74 from various regression equations (see text for further explanation).

Item Number	Type of Equations	Predictors				P-Score	Comparisons and t-Tests		
		PE	TJ	LFM	SIN COS CONT		Compared to OPNL PE + TJ	Compared to Small Sample PE	Compared to Small Sample PE+TJ+LFM
1	Operational	X	X			.1890	(comparison set)		
2	Operational	MOD	X			.1885	better		
3	Operational		X	MOD		.1996	worse**		
4	Small Sample	X				.1907	worse	(comparison set)	
5	Small Sample		X			.2145	worse**		
6	Small Sample			X		.1982	worse**	worse**	
7	Small Sample	X	X			.1947	worse*	worse**	
8	Small Sample	X		X		.1873	better	better	
9	Small Sample	X	X		X	.1890	same		
10	Small Sample	X		X	X	.1859	better	better	
11	Small Sample	X	X	X		.1884	better		(comparison set)
12	Small Sample	X	X	X	X	.1858	better		better**
13	Large Sample	X				.1921	worse**		
14	Large Sample	X			X	.1914	worse**		

Table 2. Verification on 149-case independent data sample of forecasts from summer 1974 from various regression equations (see text for further explanation).

Item Number	Type of Equations	Predictors				P-Score	Comparisons and t-Tests		
		PE	TJ	LFM	SIN COS CONT		Compared to OPNL PE + TJ	Compared to Small Sample PE	Compared to Small Sample PE+TJ+LFM
1	Operational	X	X			.2122	(comparison set)		
2		MOD	X			.2130	worse*		
3			X	MOD		.2126	worse		
4	Small Sample	X				.2347	worse**	(comparison set)	
5			X			.2345	worse**		
6				X		.2210	worse**	better**	
7		X	X			.2326	worse**	better	
8		X		X		.2257	worse**	better**	
9		X	X		X	.2261	worse**		
10		X		X	X	.2228	worse**	better**	
11		X	X	X		.2241	worse**		(comparison set)
12		X	X	X	X	.2203	worse**		better**
13	Large Sample	X				.2151	worse**		
14		X			X	.2123	worse		

Table 3--The 100 binary predictors screened for first period PoP forecasting (12-24 hr) for the winter season 1973-74. A binary variable has the value of one if the original variable from which it is derived has a value less than or equal to the binary limit; otherwise, it is zero. E-6 means $X10^{-6}$. Smoothing over n points means that, before interpolation to stations, each grid point is given the value of the mean of that point and n-1 surrounding points.

Variable	Models	Smoothing (Points)	Projection (Hours)	Binary Limits	Units
850 Height	PE	5	12	1457, 1482, 1507, 1532	M
Precipitable Water	PE	5	18	5.08, 10.16, 15.24, 20.31	Kg/m^2
Mean Relative Humidity	TJ	5	24	35, 40, 45, 50, 55, 60, 65, 70, 75, 80	Percent
12-hr Precipitation Amount	TJ	5	24	127E-7, 254E-7, 508E-7, 1016E-7, 1524E-7	M
850 Height	PE	5	24	1457, 1482, 1507, 1532	M
Boundary Layer U-wind	PE	5	24	-4, -2, 0, 2, 4, 6	M/Sec
Boundary Layer V-wind	PE	5	24	-4, -2, 0, 2, 4, 6	M/Sec
Mean Relative Humidity	PE	5	24	50, 55, 60, 65, 70, 75, 80, 85	Percent
Mean Relative Humidity	PE	9	24	50, 55, 60, 65, 70, 75, 80, 85	Percent
Boundary Layer Humidity	PE	5	24	50, 55, 60, 65, 70, 75, 80, 85	Percent
Boundary Layer Humidity	PE	9	24	50, 55, 60, 65, 70, 75, 80, 85	Percent
Second Layer Humidity	PE	5	24	50, 55, 60, 65, 70, 75, 80, 85	Percent
850 Vertical Velocity	PE	5	24	-.0005, -.0002, .0002, .0005	Mb/Sec
650 Vertical Velocity	PE	5	24	-.0005, -.0002, .0002, .0005	Mb/Sec
Total Tatals Index	PE, TJ	5	24	35, 40, 42, 44, 46, 48	-
K Index	PE, TJ	5	24	5, 10, 15, 20	-
Precipitable Water	PE	5	30	10.16, 15.24, 20.31	Kg/m^2

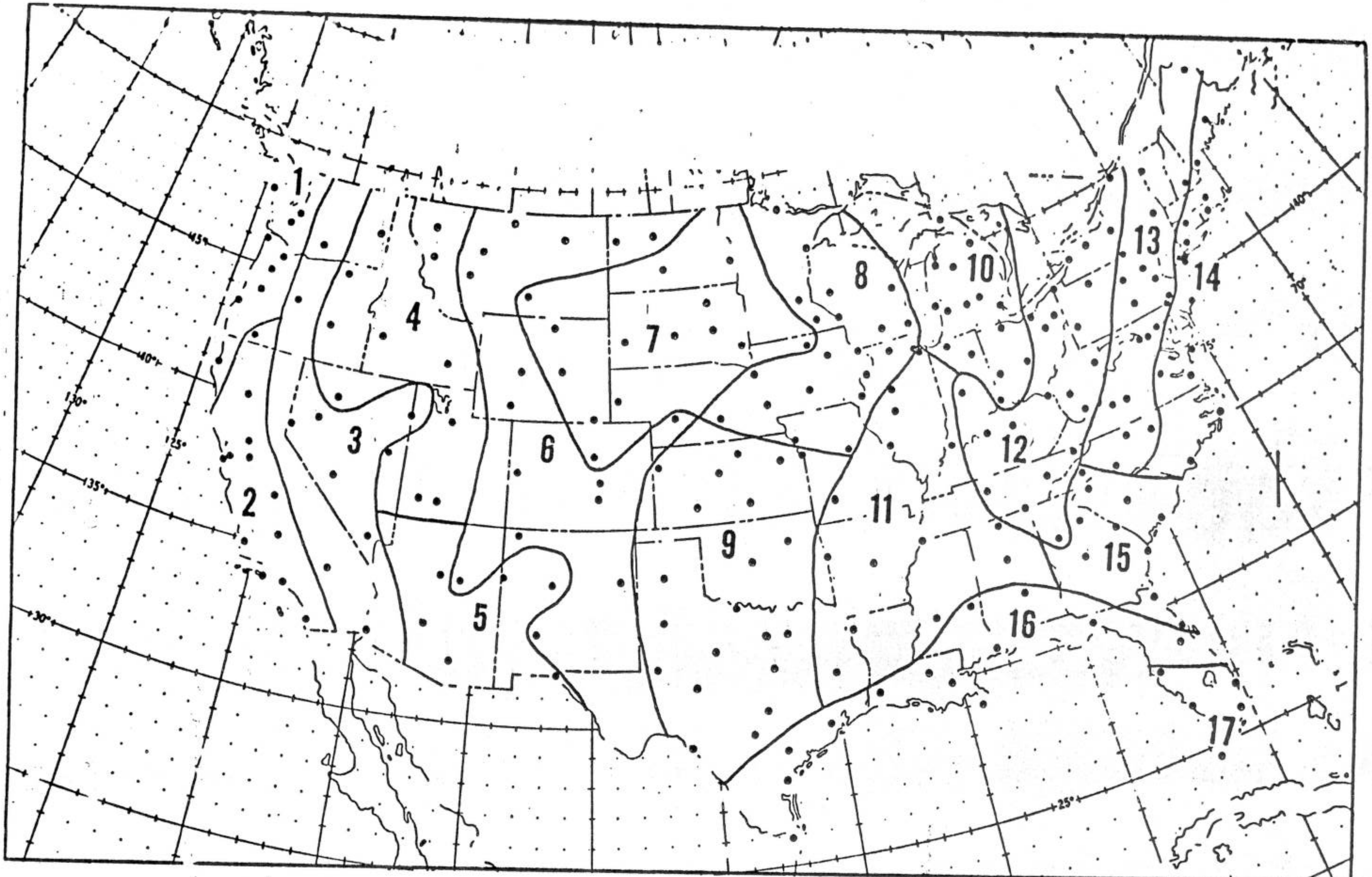

Figure 1--The 17 regions used for the operational PoP system during the 1973-74 winter season. The dots represent the 234 stations used in the development sample.

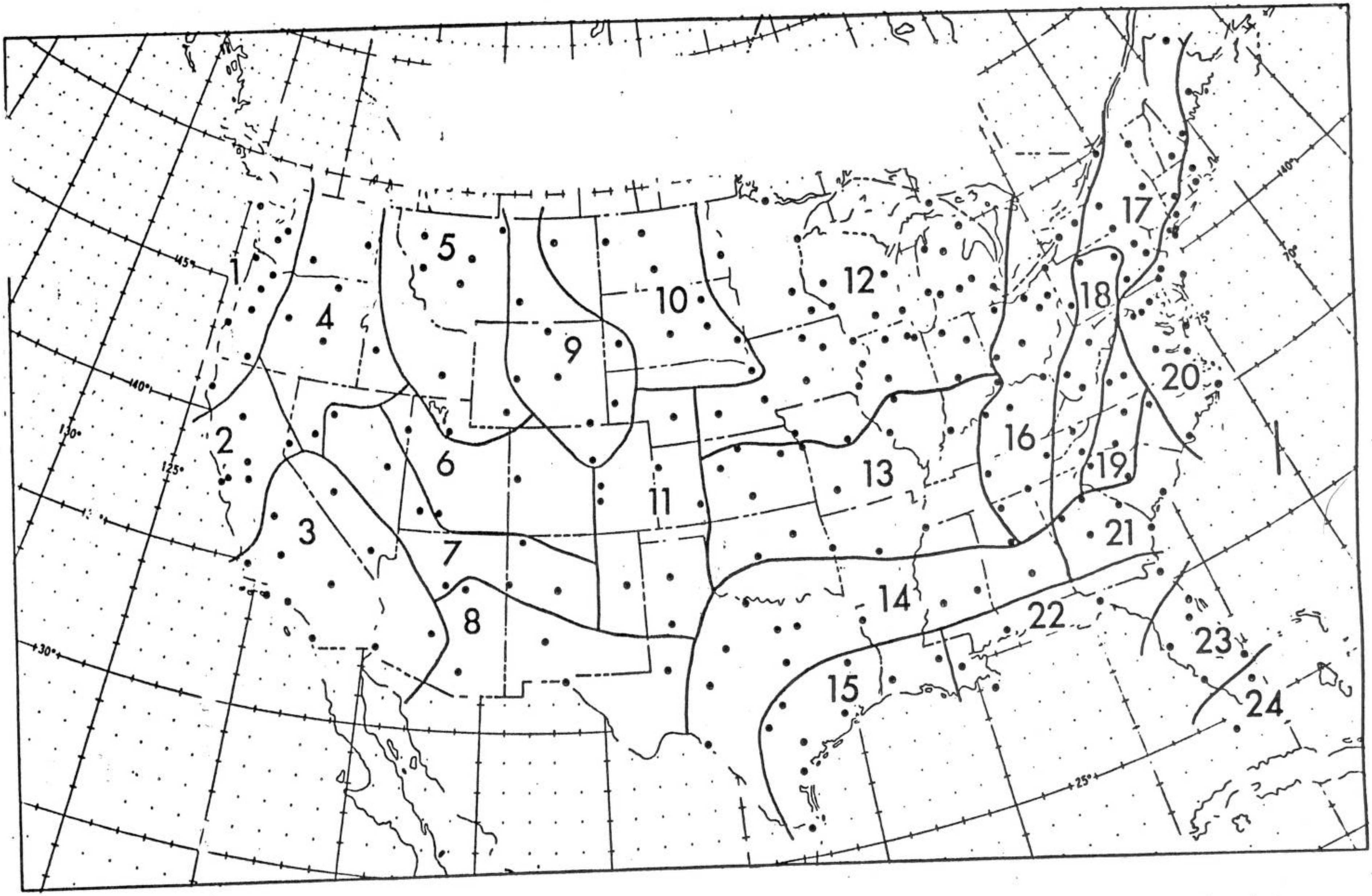

Figure 2--The 24 regions used for the operational PoP system during the summer season of 1974. The dots represent the 234 stations used in the development sample.

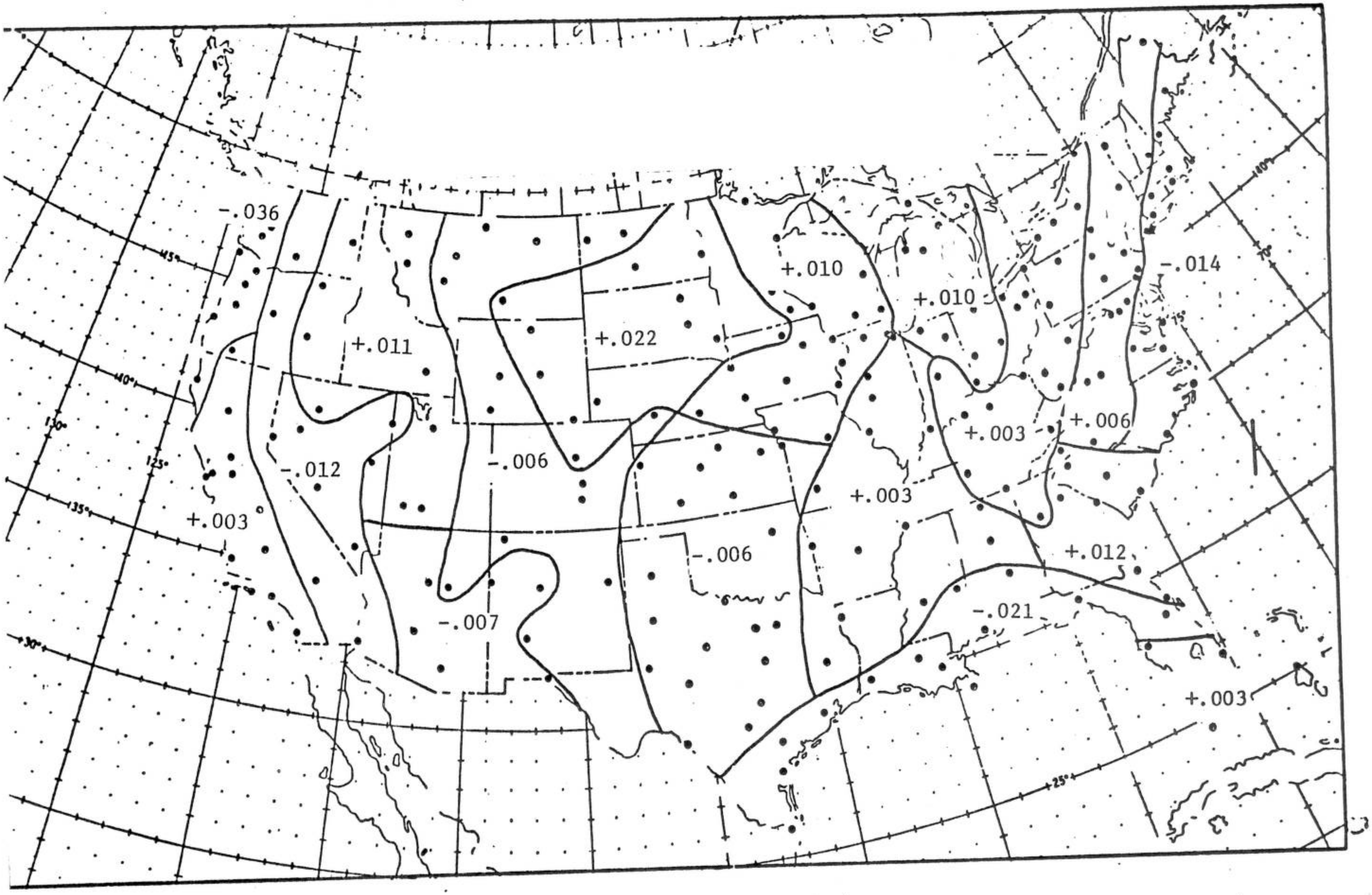

Figure 3. Values of (Item 1 P-score - Item 11 P-score). Positive values indicate Item 11 had a better (lower) P-score than Item 1. Independent data (79 days) from the winter 1973-74.

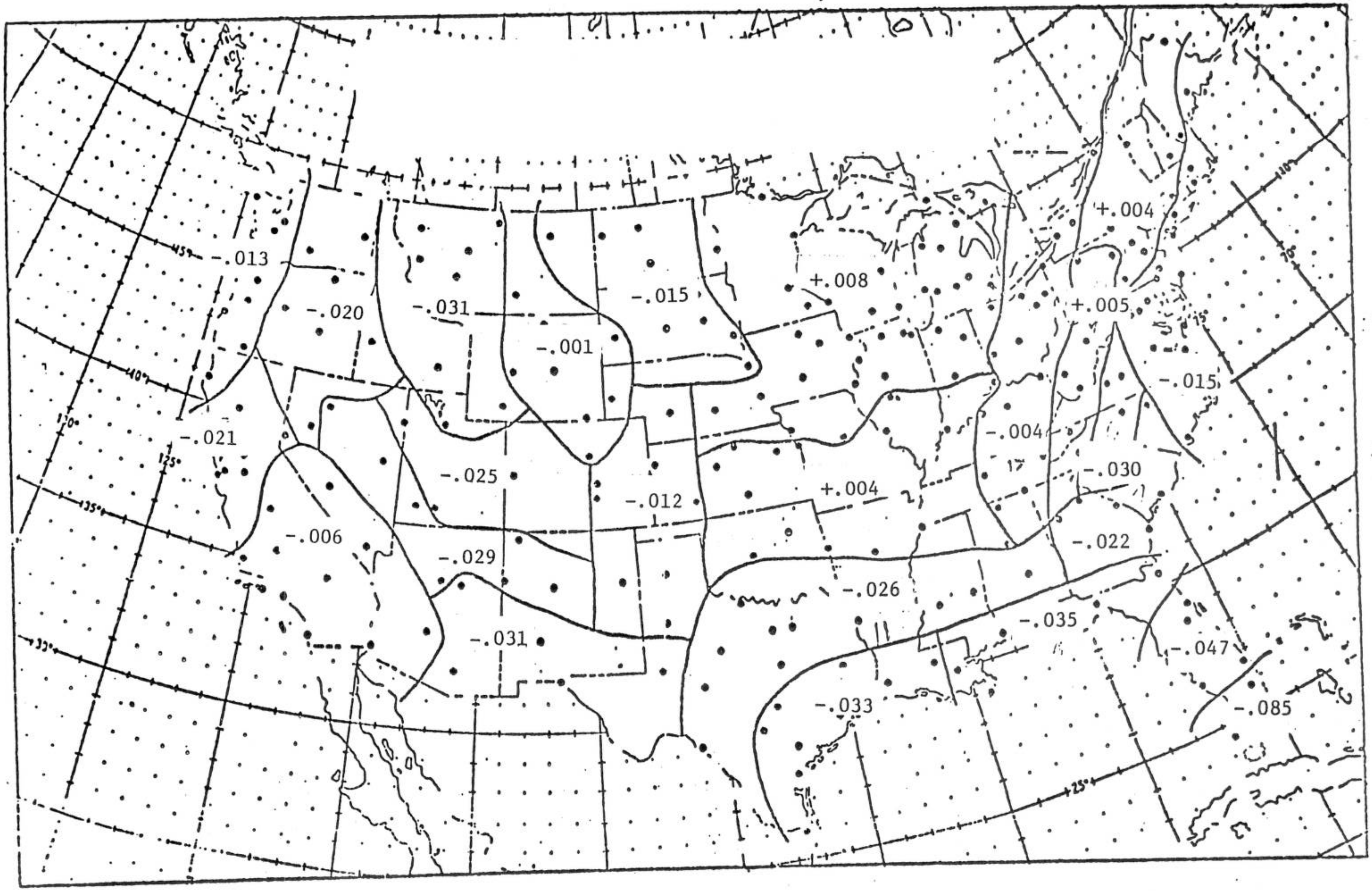

Figure 4. Same as Fig. 3 except for the 149-day independent sample from summer 1974.